



YouTube

Group14

M084020005廖冠威、M084020024黃偉豪
M084020045沈詩翰、M084020052陳柏辰
M084020056徐淳郁、M094020025劉有耘
M094020063王弘銘



01 Research Background

Background

- 成為靠YouTube維生的YouTuber
- 一年內達到4000小時觀看量 & 1000訂閱
- 分析發燒影片的關鍵因素
- 建立預測影片效益的模型





02 Dataset Introduction

Dataset Content

01

資料來源

Kaggle

02

資料集大小

40949筆資料

16個欄位

03

時間

2017/11/14~

2018/06/14

Field

欄位名稱			
Video_id	影片ID	Likes	喜歡數
Trending_date	發燒日期	dislikes	倒讚數
title	影片標題	Comment_count	評論數
Channel_title	頻道標題	Thumbnail_link	影片縮圖連結
Category_id	類別ID	Comments_disabled	是否允許評論
Publish_time	影片發佈時間	Rating_disabled	是否允許評分
tags	標籤	Video_error_or_removed	影片錯誤或移除
views	觀看數	description	影片描述

Analysis Process

Step 1 資料清理

- 新增每部影片上幾天熱門的欄位
- 獨立影片最終資料，避免重複統計。
- 新增類別資料框，針對類別作分析

Step 4 多變數分析

- 新各類別喜歡、不喜歡、評論走勢
- 喜歡比與評論比

Step 2 單變數分析

- 熱門影片類別分析
- 影片總觀看數分析
- 觀眾參與度分析

Step 3 雙變數分析

- 相互交叉分析相關性
Likes/Dislikes/comment
count/views/trending
days

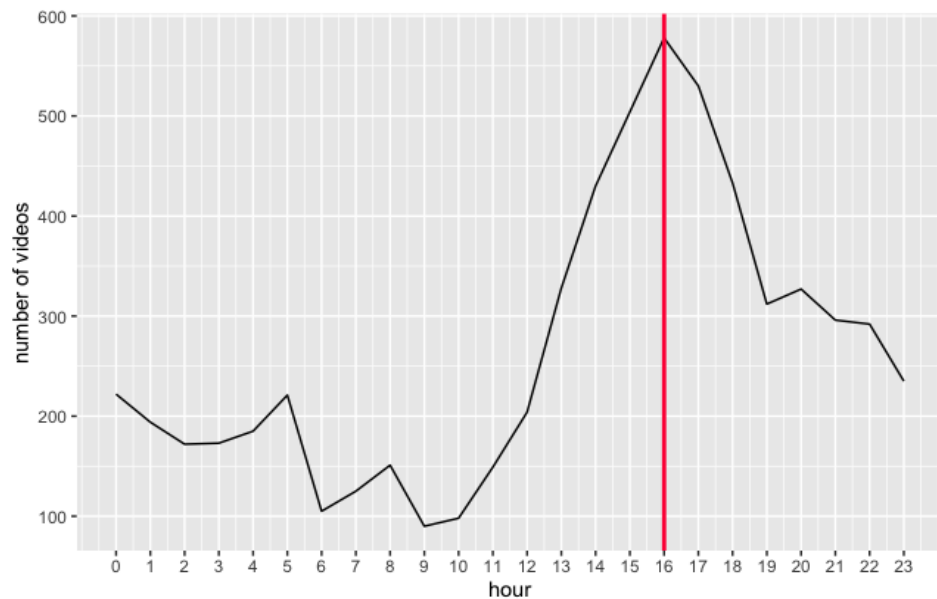




03

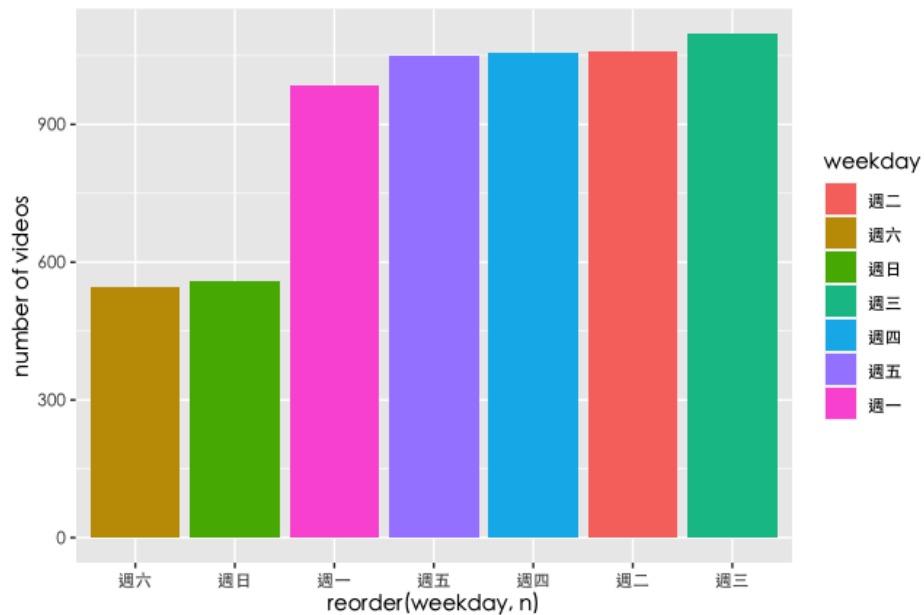
Result

Time Series Analysis(1/4)



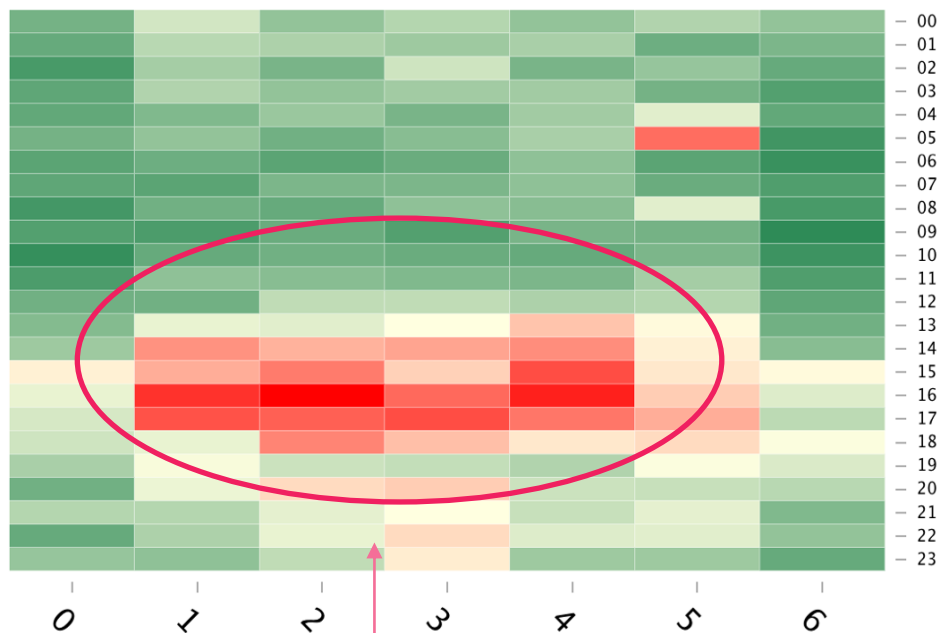
下午4點為影片發布熱點

Time Series Analysis(2/4)



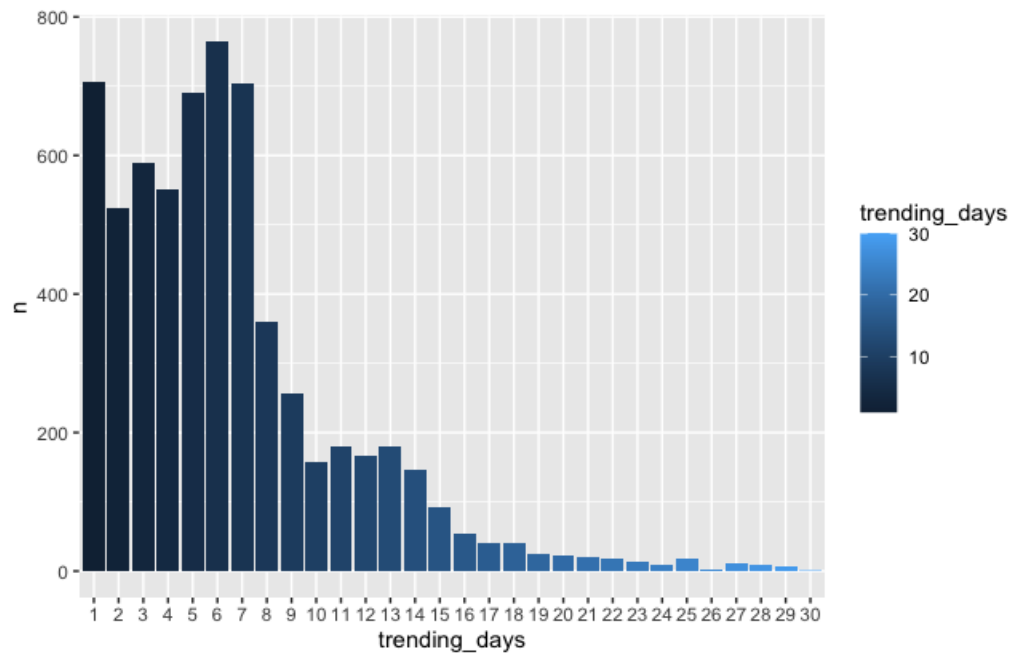
週三、週二影片發布量最多

Time Series Analysis(3/4)



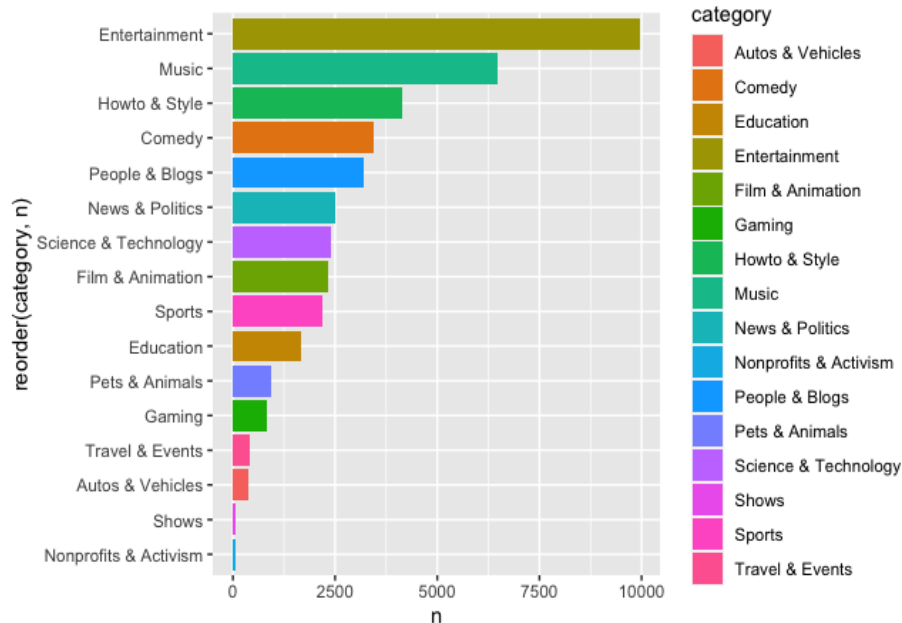
每周二到四下午四點左右為影片發布熱點

Time Series Analysis(4/4)



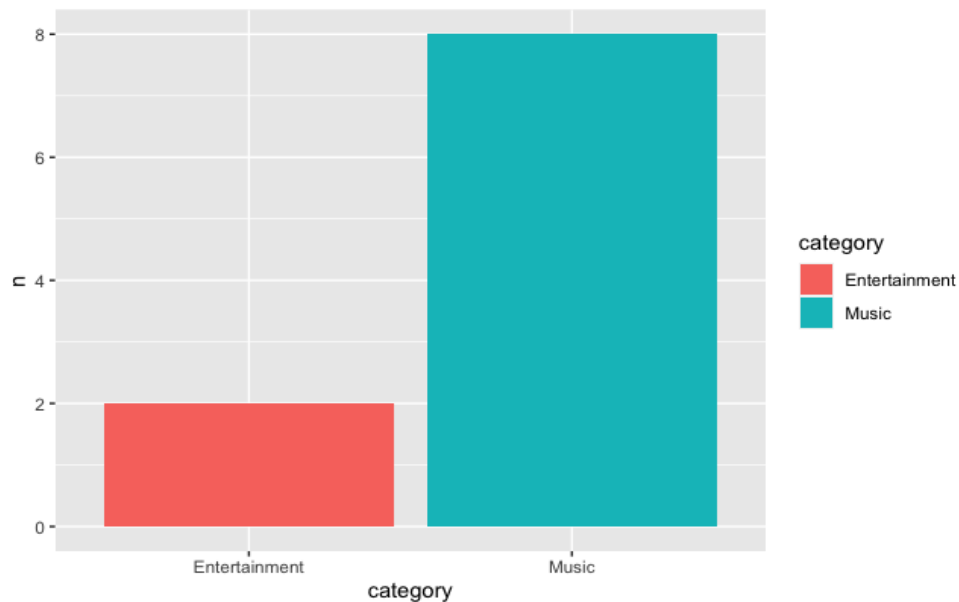
影片上熱門的天數集中在1-7天

Most popular video category (1/2)



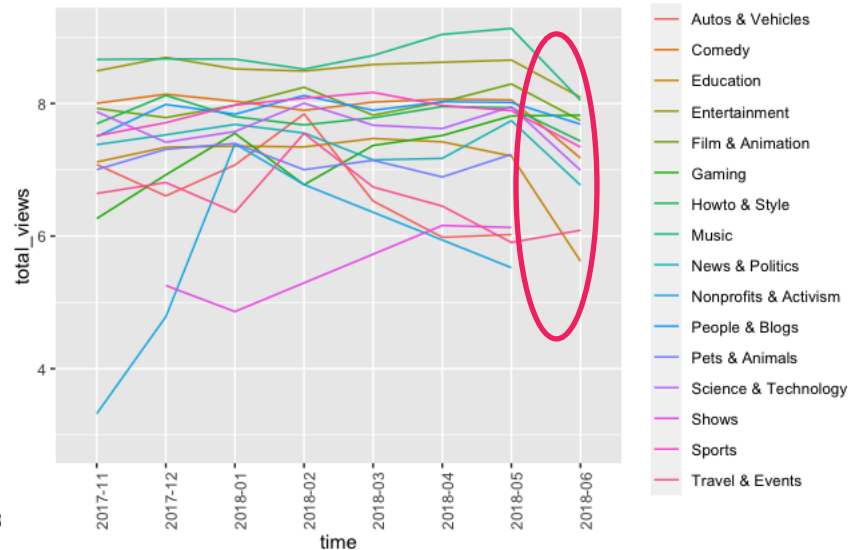
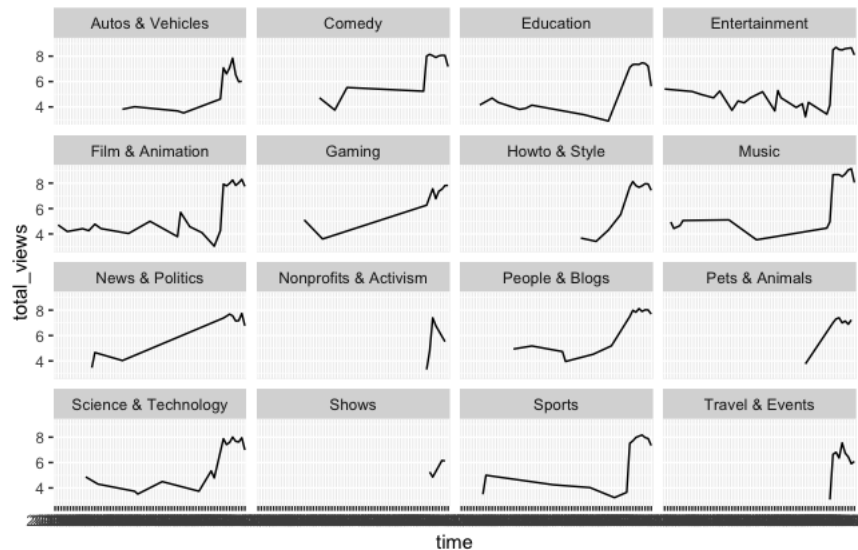
Entertainment 為上熱門次數最多的類別

Most popular video category(2/2)



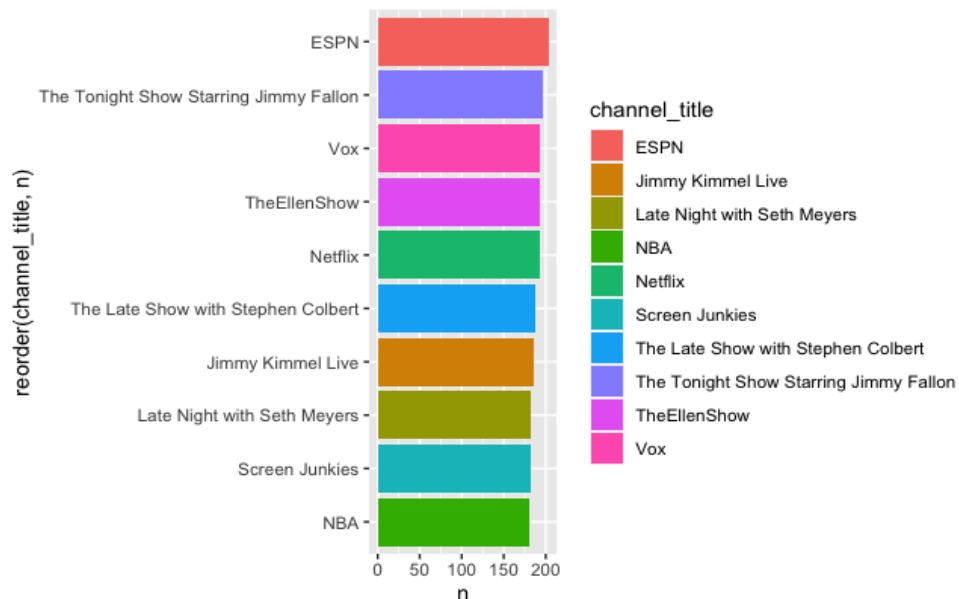
娛樂 2 : 8 音樂

Most potential video category



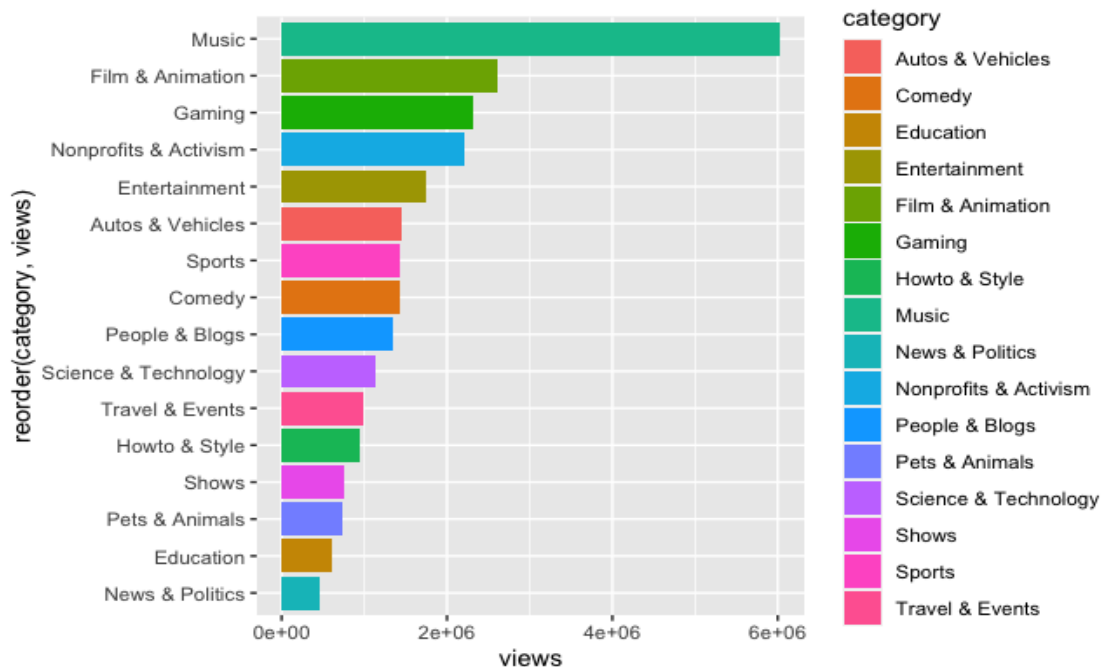
各類別影片觀看次數成長趨勢

Most popular channels



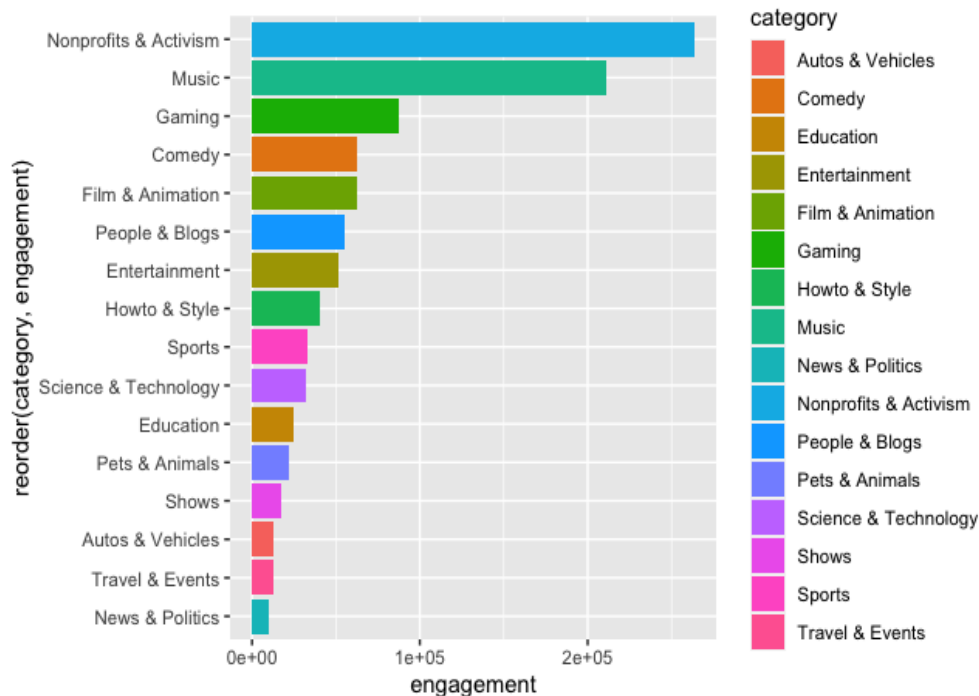
前十個最常上熱門的頻道

Highest Video Views



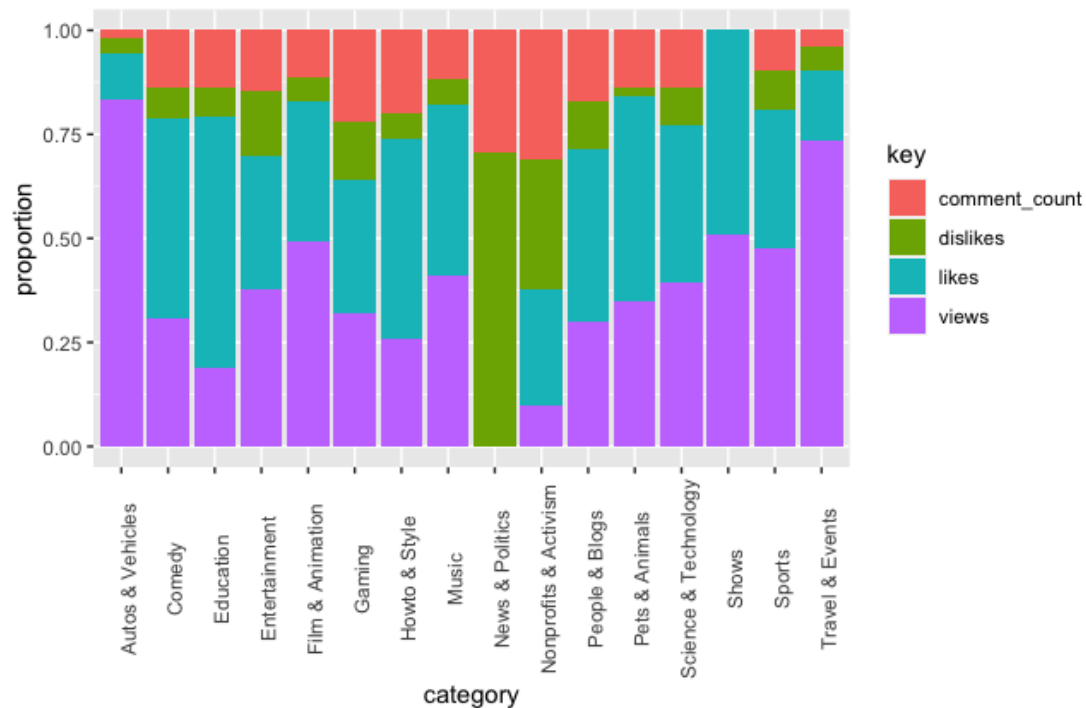
Music 為擁有最高總觀看次數的種類

Highest Audience participation (1/2)

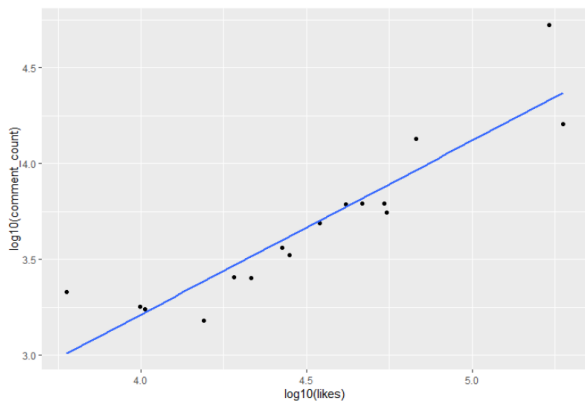


Nonprofits & Activism 為擁有最高觀眾參與度的種類

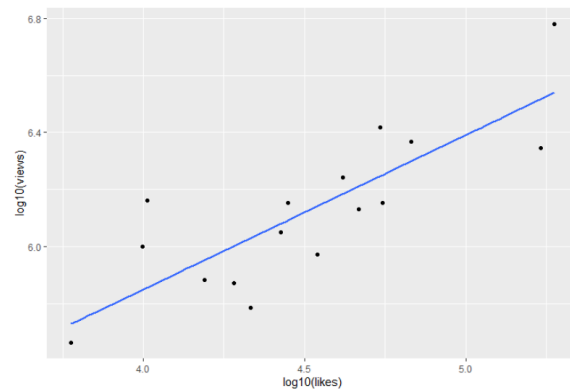
Highest Audience participation (2/2)



Correlation analysis-Likes



```
data: C$likes and C$comment_count
t = 5.0502, df = 14, p-value = 0.0001772
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5115436 0.9291330
sample estimates:
      cor
0.8035012
```

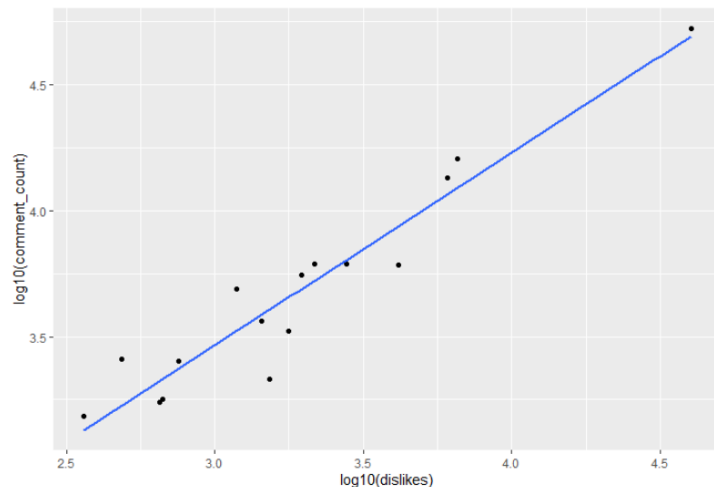


```
data: C$likes and C$views
t = 5.6096, df = 14, p-value = 6.436e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5721658 0.9399860
sample estimates:
      cor
0.8319179
```

P-value<0.05且cor>0.8 兩著有高度正相關

Likes 和 Views 、 Comment count 有顯著正相關

Correlation analysis-Dislikes

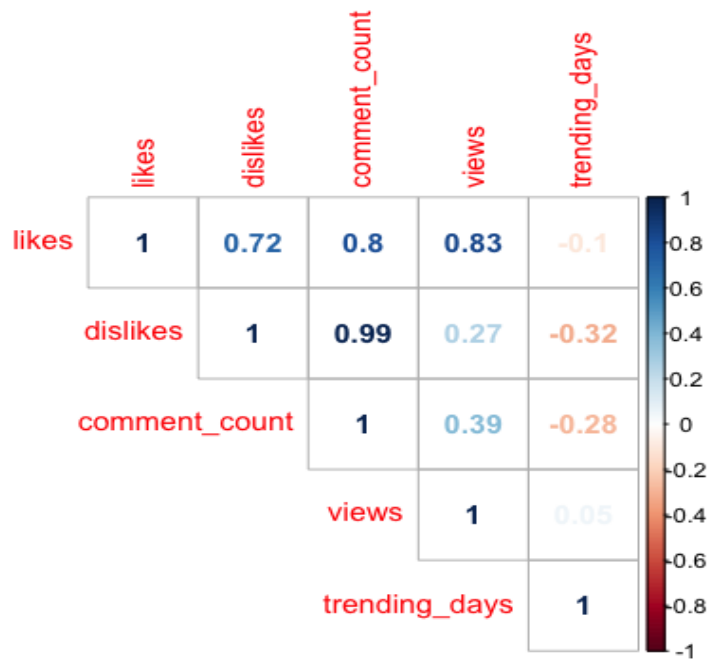


```
data: C$dislikes and C$comment_count
t = 23.628, df = 14, p-value = 1.11e-12
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9639327 0.9958333
sample estimates:
      cor
0.9876924
```

P-value<0.05且cor>0.8 兩著有高度正相關

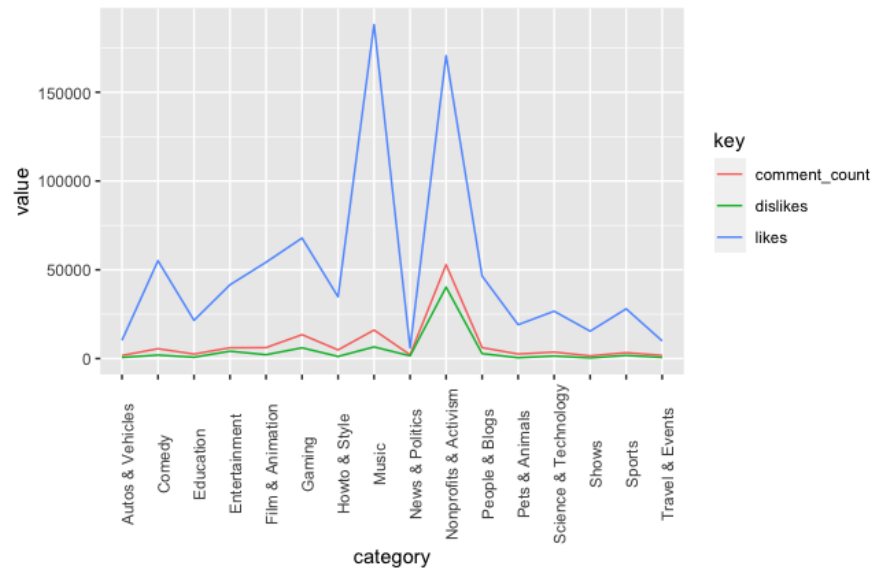
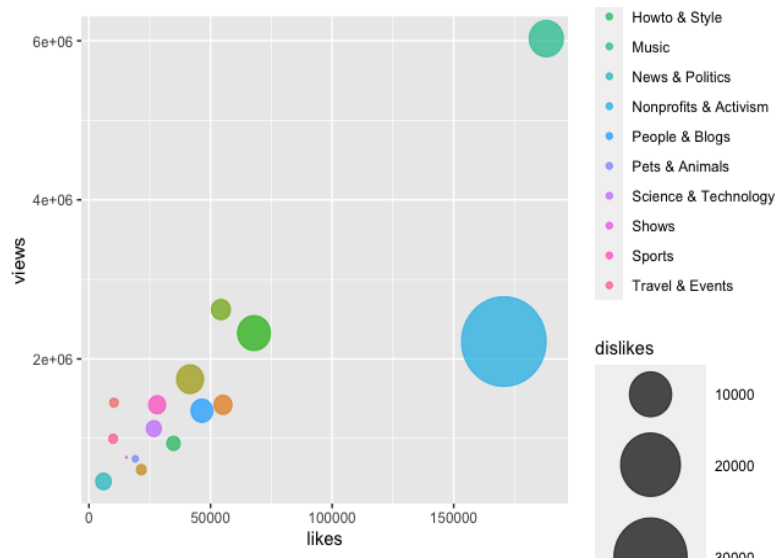
Dislikes 和 **Comment count** 有顯著正相關

Trending (1/4)



各類別 Likes\Dislikes\comment cout 走勢

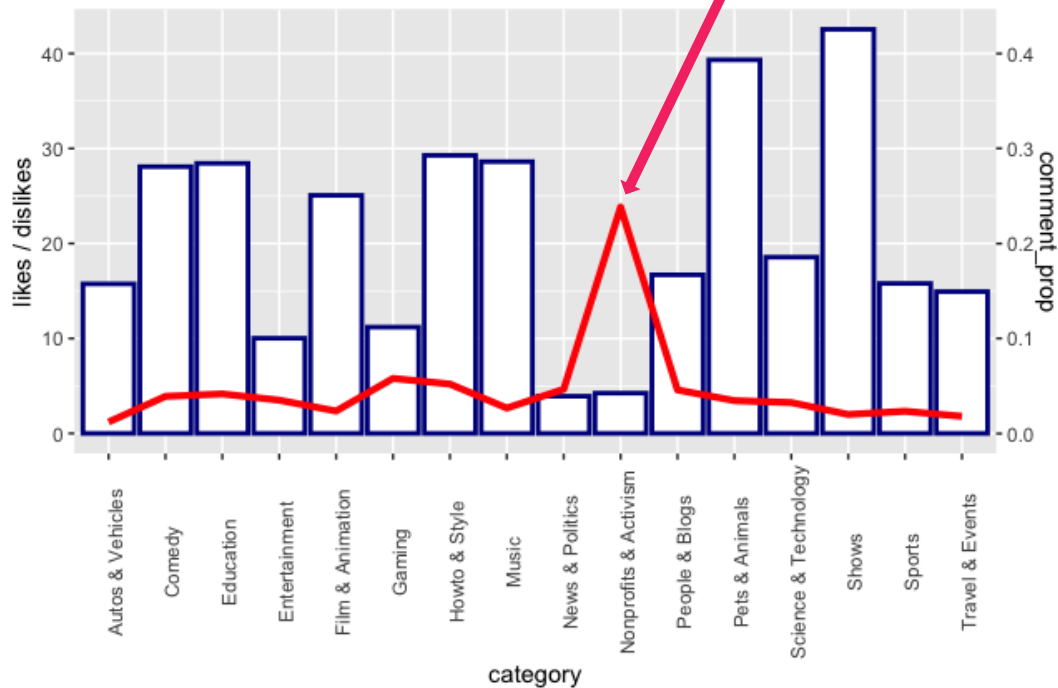
Trending (2/4)



若 Dislikes 多，即使 Likes 多卻不會提升 Views

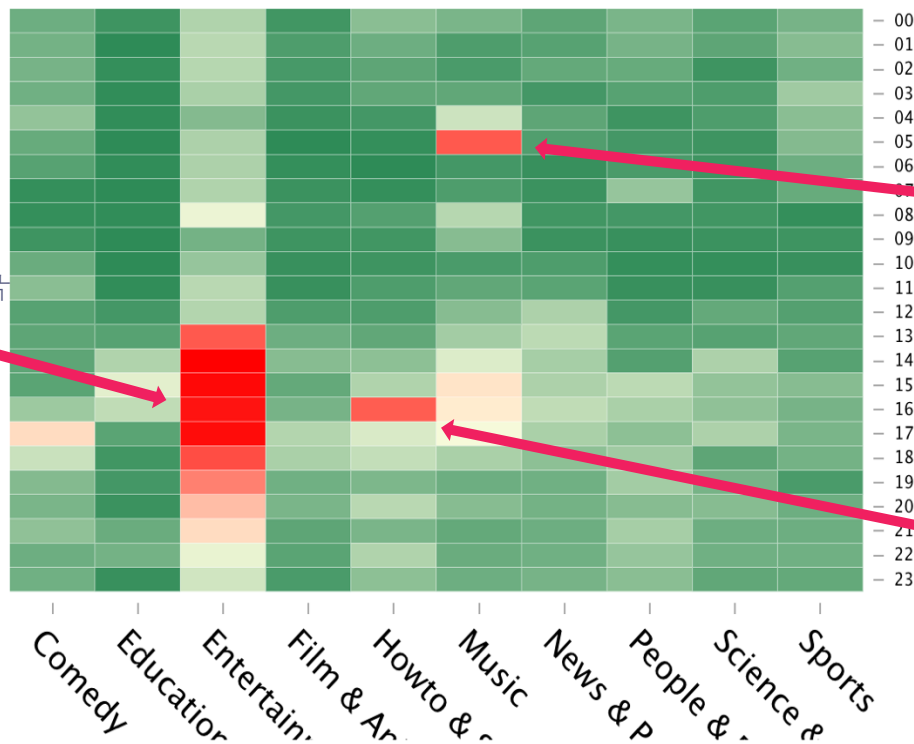
Trending (3/4)

Nonprofits & Activism 喜歡
比低但評論比高



Trending (4/4)

此時發布 **Entertainment** 類型影片
效益最大



此時發布 **Music** 類型
影片效益最大

此時發布 **Howto & Style** 類型
影片效益最大



04

Future Research

Research Content



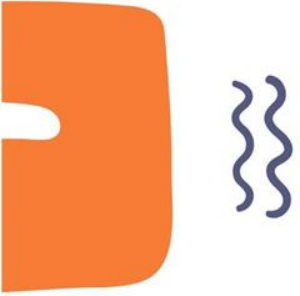
1.

- 加入台灣資料
- 比較兩國文化差異
- 困難點：時間較短資料數量少

2.

- 將收入衡量指標加入預測。
(CPM、RPM)
- 困難點：缺少衡量項目之一的
「觀眾群地區」





Thanks!

